X-514-71-312

NASA TM X- 65663

ATTITUDE DETERMINATION AND SENSOR ALIGNMENT VIA WEIGHTED LEAST SQUARES AFFINE TRANSFORMATIONS

PAUL B. DAVENPORT

AUGUST 1971

GSFC

GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

AAS NO. 71-396

ATTITUDE DETERMINATION AND SENSOR ALIGNMENT VIA WEIGHTED LEAST SQUARES AFFINE TRANSFORMATIONS

Paul B. Davenport

GODDARD SPACE FLIGHT CENTER

Greenbelt, Maryland

AAS/AIAA
ASTRODYNAMICS SPECIALISTS CONFERENCE 1971
FT. LAUDERDALE, FLORIDA
AUGUST 17 - 19 1971

ATTITUDE DETERMINATION AND SENSOR ALIGNMENT VIA WEIGHTED LEAST SQUARES AFFINE TRANSFORMATIONS

Paul B. Davenport
Goddard Space Flight Center
Greenbelt, Maryland

ABSTRACT

The position and orientation of a vehicle, sensor, actuator, etc., relative to a fixed set of axes (another vehicle, sensor, etc.) may be represented mathematically by an affine transformation (linear transformation plus a translation). Symbolically, $Z = MX + V$, where M is an mxn matrix, V and Z are mx1 column vectors, and X is an nx1 vector. This matrix equation also represents any linear relationship between known inputs X and measured outputs Z; e.g., if X is the first n powers of a scalar x then each component of Z is a polynomial in x. The weighted least squares estimate of M and V is discussed assuming that various measurements Z are given (along with the input X). Although there are m(n+1) parameters to be estimated, a simple weighting function allows a solution by inverting only an nxn matrix. This case, including constraints on M (orthogonal, rotation, symmetric and skew-symmetric) will be examined in detail.

CONTENTS

	Page
ABSTRACT	iii
MOTIVATION	1
THE MODEL	2
THE CONDITION EQUATIONS	4
THE CASE OF AN EXPLICIT MATRIX SOLUTION	7
SEQUENTIAL SOLUTIONS	8
A SINGULAR CASE	10
THE GENERAL SOLUTION	12
CONSTRAINTS	14
ORTHOGONAL AND ROTATIONAL SOLUTIONS	16
ATTITUDE DETERMINATION	20
SYMMETRIC AND SKEW-SYMMETRIC SOLUTIONS	27
REFERENCES	29
BIBLIOGRAPHY	31

ATTITUDE DETERMINATION AND SENSOR ALIGNMENT VIA WEIGHTED LEAST SQUARES AFFINE TRANSFORMATIONS

MOTIVATION

Many problems of navigation and guidance (sea, air, space) can be stated fundamentally as the determination of the orientation and/or position of an axis or axes relative to another set of axes (e.g., coordinate axes). The very heart of navigation is to determine the attitude and/or position of a vehicle so that the vehicle may be headed toward (or held into) a desired position and/or orientation. To perform these tasks numerous sensors and actuators are usually employed. This then requires the knowledge of the position and orientation of each device relative to the vehicle axes. Thus, there is the dual problem of determining the location of each instrument axes relative to the vehicle axes so that the orientation and/or position of the vehicle axes relative to some underlying coordinate axes may be ascertained.

The first problem (instrument alignment and calibration) is usually performed once or rather infrequently under laboratory conditions whereas the attitude and/or position evaluation is performed in the "field" by the navigator (human or computer). For precise missions of long duration the "navigator" may be required to perform instrument alignment and calibration calculations in addition to the attitude and/or position determinations. Gyro drift, thermal bending, stresses, fatigue, etc. may combine to produce unacceptable tolerances. In this event, it would be convenient if the navigator's prior skills (computer algorithms) for determining orientation and/or position could be applied to the

alignment problem. Here, a mathematical model and techniques will be presented which apply to a large class of alignment and calibration problems, attitude determination, and position determination. The development is via matrix algebra so the results apply to any vector space of arbitrary finite dimension.

THE MODEL

Regardless of the instrumentation, physical vector quantities are generally being sensed — either directions (star, sun, horizon, etc.) or multiple scalar quantities which collectively define a vector (triads of magnetometers, gyros, accelerometers, etc.). Thus, the output of many sensors or actuators can be represented by mx1 matrices (vectors) of components (or direction cosines) in an m-dimensional Euclidean space. Such m-tuples will be denoted as Z. Likewise, the components of the physical vector (magnetic field, sun direction, etc.) are known relative to some underlying coordinate system. These vectors will be denoted by the vector X. The generalized problem of navigation could then be stated as the determination of the relationship between X and the measurement vector Z. For most practical systems this functional relationship can be expressed as

$$Z = MX + V \tag{1}$$

where M is an mxn constant matrix and V is an mx1 fixed matrix (vector).* X is an nx1 vector which may be just the components of the physical vector being measured (in which case $m = n$) or more generally X may be an n-tuple whose entries are any known functions of the components, such as powers of

*No distinction is made in notation between vectors and matrices except that the latter part of the alphabet (starting with T) will be reserved for vectors.

the components. As a special case X might be a scalar and M the coefficients of a polynomial.

The above relationship is known as an affine transformation (linear transformation plus a translation). It includes as special cases; linear transformations, $V = 0$; and translations, $M = I$. If M is non-singular, then it has the geometric interpretation of defining the orientation of one coordinate system relative to another. This includes a rotation of axes, reflections, non-orthogonal axes (shearing), and a different scale factor for each component. In some applications additional constraints may be placed upon M and V. For example, attitude determination (orientation of rigid body with one point fixed) requires M to be a rotation matrix (orthogonal with determinant of plus one) and V be zero. Equation (1) also models a system of m single axis devices, each with different scale factors, located non-orthogonally from the center of the vehicle. Furthermore, there are no restrictions of smallness, i.e., the displacements may be large.

When M, V, and X are given it is trivial to compute Z (the coordinates relative to a vehicle, a conglomerate of instruments, etc.) so as to point a sensor or actuator as desired. A more pertinent problem to the navigator, however, is to determine M and/or V given the local measurements Z (containing errors) and the vectors X. In this case, (1) can be considered as a system of linear equations containing $m(n + 1)$ unknowns (the elements of M and V). Thus, if $n + 1$ independent vectors X_k and the corresponding measurements Z_k (each with m independent components) are known M and V are determined uniquely. This is not generally the case, however; one usually has insufficient data or an over-determined system with inconsistent equations due to

measurement errors. The classical approach to this dilemma is to seek an "estimate" of the parameters which is best in some sense. Here, the estimate which minimizes the weighted sums of squares of the vector norm will be discussed, i.e., the minimization of

$$f(M,V) = \sum_{k=1}^{\ell} [Z_k - (MX_k + V)]^T P_k [Z_k - (MX_k + V)] \tag{2}$$

where the P_k are positive definite symmetric weight matrices (e.g., the variance-covariance of Z_k) indicating the relative accuracy of Z_k. The different Z_k may represent measurements from different types of instruments or readings from the same instrument at different times. The summation is taken over all such measurements ℓ. Expanding (2), bearing in mind that $P_k^T = P_k$, and collecting terms gives:

$$f(M,V) = \sum [Z_k^T P_k Z_k - 2Z_k^T P_k MX_k + (MX_k)^T P_k MX_k$$
$$+ V^T P_k V - 2Z_k^T P_k V + 2V^T P_k MX_k] . \tag{3}$$

The summation indices are hereafter omitted for convenience. Unless otherwise stated the summation ranges over all measurements.

THE CONDITION EQUATIONS

The necessary conditions that f(M,V) have a minimum are

$$\frac{\partial f}{\partial m_{ij}} = 0 ,$$
$$\frac{\partial f}{\partial v_i} = 0 , \tag{4}$$

for all i and j ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$). Formally, this results into $m(n + 1)$ linear scalar equations in the $m(n + 1)$ unknowns (m_{ij} and v_i). A more elegant and informative procedure is to retain the matrix notation and express the resulting conditions as matrix equations rather than a large number of scalar equations. To facilitate this, the following definition is made: If h is a scalar function of an mxn matrix Q with elements q_{ij} then the "gradient" of h(Q), denoted as $\nabla h(Q)$ or simply ∇h, is the mxn matrix with elements $(\nabla h)_{ij}$ given by

$$(\nabla h)_{ij} = \frac{\partial h}{\partial q_{ij}} .$$

The gradient of several elementary scalar functions (which are adequate for the present discussion) are given below:

$\nabla h(U) = W$, if $h(U) = W^T U = U^T W$, W and U nx1 matrices

$\nabla h(Y) = (P^T + P)Y$, if $h(Y) = Y^T P Y$, P mxm, Y mx1

$\nabla h(N) = YU^T$, if $h(N) = Y^T N U$, N mxn

$\nabla h(N) = (P^T + P)NUU^T$, if $h(N) = (NU)^T PNU$.

These identities follow directly from the definition and the rules of matrix multiplication. For a function of several matrices, e.g., h(P,Q) the notation $\nabla(P)h$ denotes the gradient of h with respect to P only. In other words, $\nabla(P)$ operates on h as if h were a function of P alone.

In terms of the above definitions, the condition Equations (4) can be written as

$$\nabla(M) f = 0 ,$$

$$\nabla(V) f = 0 .$$

Performing the indicated operators on (3) and equating to zero yields the following simultaneous matrix equations:

$$\sum P_k M X_k X_k^T - \sum P_k Z_k X_k^T + \sum P_k V X_k^T = 0$$

$$\sum P_k V - \sum P_k Z_k + \sum P_k M X_k = 0 . \tag{5}$$

The Equations (5) still represent m(n + 1) linear equations, but offer a notational advantage over those implied by (4) in that M and V appear explicitly rather than their components. Apparently, generalized techniques for solving such systems are non-existent except to re-write the equations as a single matrix equation of the form CY = W where Y and W are column vectors of dimension m(n + 1). A convenient notation exists for accomplishing this, but the numerical solution may present a prodigious amount of computation even for relatively small n and m.

This last representation of the problem (CY = W) could have also been obtained by using classical linear estimation results rather than the above approach. (1) can be re-written for each observation as $Z_k = C_k Y$, where the matrix C_k is a function of X_k and Y is a column vector composed of the unknowns m_{ij} and v_i in some order (e.g., the columns of M plus V concatenated). If the classical least squares conditions are applied to the matrix equation representing all such observations (assuming each vector observation is independent) then one obtains the same matrix equation as that derived from re-writing (5) in the form CY = W. It is not the intent here to do this explicitly, but to appeal to well-known least squares results to insure that a solution to (5) exists. The development leading to (5) was selected since it gives more insight into the nature of the solution and is more amenable to constraints which will be considered later.

THE CASE OF AN EXPLICIT MATRIX SOLUTION

The complexity of (5) can be attributed to the generalized statistical model rather than the geometrical model. If the weight matrices are of the form $P_k = p_k P$ for all k (P a fixed positive definite matrix and p_k a positive scalar) then (5) may be solved explicitly in a closed simple form. Under these conditions (5) reduces to:

$$\begin{aligned} P M A_0 + P V X_0^T &= P B_0 \\ P M X_0 + s P V &= P Z_0 , \end{aligned} \tag{6}$$

where

$$\begin{aligned} A_0 &= \sum p_k X_k X_k^T , \quad B_0 = \sum p_k Z_k X_k^T , \\ X_0 &= \sum p_k X_k , \quad Z_0 = \sum p_k Z_k , \end{aligned} \tag{7}$$

and

$$s = \sum p_k .$$

Since P was defined to be non-singular (6) is equivalent to

$$\begin{aligned} M A &= B , \\ V &= \frac{1}{s} (Z_0 - M X_0) , \end{aligned} \tag{8}$$

where

$$A = A_0 - \frac{1}{s} X_0 X_0^T , \quad B = B_0 - \frac{1}{s} Z_0 X_0^T .$$

Hence, if $P_k = p_k P$, the questions of existence, uniqueness, and the solution itself depend only on the simple matrix equation $MA = B$ where A is a constant nxn symmetric matrix and B is a fixed mxn matrix.

If A is non-singular, then (8) has a unique solution

$$M = B A^{-1}$$

$$V = \frac{1}{s} (Z_0 - M X_0)$$

which minimizes (2) (the sufficiency of (8) for a minimum follows from the nature of the function and linear least squares theory). The least squares translation is given by $M = I$ (the identity matrix), $V = \frac{1}{s} Z_0$; whereas the least squares linear transformation is $M = B_0 A_0^{-1}$ and $V = 0$.

SEQUENTIAL SOLUTIONS

A familiar identity from recursive least squares

$$\left(C - \frac{1}{a} U W^T\right)^{-1} = C^{-1} (I - b U W^T C^{-1}), \tag{9}$$

with

$$\frac{1}{b} = W^T C^{-1} U - a ,$$

may be employed to examine three different models (full affine, translation only, and linear transformation only) with a single matrix inversion. If V' denotes the vector of the least squares translation, L the matrix of the least squares linear transformation, M and V the matrix and vector of the least squares affine transformation, then

$$V' = \frac{1}{s} Z_0 ,$$

$$L = B_0 A_0^{-1} ,$$

$$\frac{1}{r} = X_0^T A_0^{-1} X_0 - s ,$$

$$A^{-1} = A_0^{-1} (I - r X_0 X_0^T A_0^{-1}) ,$$

$$M = B A^{-1} ,$$

$$= L - r (L X_0 - Z_0) X_0^T A_0^{-1} ,$$

$$V = \frac{1}{s} (Z_0 - M X_0) .$$

The above formulas imply the existence of A_0^{-1} under the earlier assumption that A was non-singular. This is indeed the case, but the converse (A^{-1} exists if A_0^{-1} exists) is not true if $X_0^T A_0^{-1} X_0 = s$.

The identity (9) may also be applied to a recursive solution of (8). If a superscript is added to each of the intermediate quantities in (7) to denote the summation limit, e.g.,

$$A_0^{\ell} = \sum_{k=1}^{\ell} p_k X_k X_k^T ,$$

then

$$A_0^{\ell} = A_0^{\ell-1} + p_{\ell} X_{\ell} X_{\ell}^T , \quad B_0^{\ell} = B_0^{\ell-1} + p_{\ell} Z_{\ell} X_{\ell}^T ,$$

$$X_0^{\ell} = X_0^{\ell-1} + p_{\ell} X_{\ell} , \quad Z_0^{\ell} = Z_0^{\ell-1} + p_{\ell} Z_{\ell} ,$$

$$s^{\ell} = s^{\ell-1} + p_{\ell} .$$

Therefore,

$$(A_0^{\ell})^{-1} = (A_0^{\ell-1})^{-1} \left[I - q X_{\ell} X_{\ell}^T (A_0^{\ell-1})^{-1} \right],$$

where

$$q = X_{\ell}^T (A_0^{\ell-1})^{-1} X_{\ell} + 1/p_{\ell} ,$$

provided $(A_0^{\ell-1})^{-1}$ exists.

A SINGULAR CASE

When the matrix A is singular, then there is insufficient data to define the model uniquely. For some applications, however, any solution which fits the data might be adequate. Since A is symmetric, there exists an orthogonal matrix Q such that $A = QDQ^{-1}$ where D is a diagonal matrix with entries d_j ($j = 1, 2, \ldots, n$). Eq. (8) can then be written as $M'D = B'$ with $M' = MQ$ and $B' = BQ$. Let r denote the number of non-zero elements of D and assume that Q and D are such that these non-zero elements occupy the first r columns (rows) of D. The above equations then give

$$m'_{ij} = \frac{1}{d_j} b'_{ij}, \quad (i = 1, 2, \ldots, m)$$

$$(j = 1, 2, \ldots, r)$$

with m'_{ij} arbitrary for $j = r + 1, \ldots, n$. The consistency of the solution can be justified by appealing to classical linear least squares estimation theory. To be consistent requires $b'_{ij} = 0$ ($j = r + 1, \ldots, n$), i.e., the last $n - r$ columns of B' are zero.

Let D^+ be a diagonal matrix with entries $d_j^+ = 1/d_j$ ($j = 1, 2, \ldots, r$) and $d_j^+ = 0$ for $j = r + 1, \ldots, n$. The above solution can then be expressed as:

$$M\,Q \;=\; M' \;=\; B'\,D^{+} + N\,,$$

where N is an mxn matrix whose first r columns are zero and the last n - r columns are arbitrary vectors. Solving for M yields

$$\begin{aligned} M &= B\,Q\,D^{+}\,Q^{-1} + N\,Q^{-1} \\ &= B\,A^{+} + N\,Q^{-1} \end{aligned}$$

where $A^{+} = Q\,D^{+}Q^{-1}$ is the pseudo-inverse of A [9]. Denoting the non-zero columns of N as $N_{r+1}, \ldots, N_n$ and the corresponding columns of Q (the eigenvectors associated with the zero eigenvalues of A) as $Q_{r+1}, \ldots, Q_n$, then $N\,Q^{-1} = \Sigma\, N_j\, Q_j^T$ (summed from r + 1 to n).

Analogous to the least squares vector of least norm, one may obtain the unique least squares matrix of least norm (norm of M defined as tr $(M^T M)$ = tr $(M\,M^T)$, tr denotes "trace of") by letting N be the null matrix. On the other hand, since M may represent a linear transformation, a solution which is "closest" to the identity transformation may be desirable, i.e., minimizes the norm of I - M when M is square. This solution is obtained by setting $N_j = Q_j$ ($j = r + 1, \ldots, n$). Both of these special solutions are easily verified by using the well-known properties of the pseudo-inverse and trace function.

The vector V is always given by the second equation of (8) whether A is non-singular or not. Thus, in the singular case

$$V \;=\; \frac{1}{s}\left[Z_0 - B\,A^{+}\,X_0 - \sum\, (Q_j^T\,X_0)\,N_j\right]$$

(summation from r + 1 to n). If $Q_j^T\,X_0$ is non-zero for some j, then the arbitrariness of N may be used to eliminate V instead of imposing the above conditions on M.

THE GENERAL SOLUTION

Thus far, the solution to the simultaneous matrix equations (5), which provides the minimum of (2), has been exhibited under all conditions for the special weighting $P_k = p_k P$. If each vector measurement Z_k is statistically independent and the variance-covariance matrices of each Z_k differ only by a multiplicative constant then these solutions provide the minimum variance solutions. This is the situation when all vector measurements relate to the same type of instrument (whose variance-covariance matrix varies only by a scalar) or when each vector measurement is composed of m independent scalar measurements with the same variance. In the general case, one might forsake a minimum variance requirement in order to obtain a simple solution by assigning a single weight to each vector. In many instances this may be an adequate solution or serve as an initial approximation for an iterate scheme.

As noted previously, the computational complexity of the problem soars when the general weight is considered. The dimension of the matrix to be inverted is increased by a factor of m which enlarges the computations by a factor of order m^3.

Should accuracy considerations dictate the additional effort, however, there is available notation for deriving the larger system of linear equations in matrix form. This is via the direct product (also tensor or Kronecker product) of matrices. If A is an mxn matrix and B a m'xn' matrix, the direct product of A and B, in that order, is defined by the partitioned matrix

$$A \otimes B = \begin{pmatrix} a_{11}\,B & a_{12}\,B & \dots & a_{1n}\,B \\ a_{21}\,B & a_{22}\,B & \dots & a_{2n}\,B \\ \vdots & & & \vdots \\ a_{m1}\,B & a_{m2}\,B & \dots & a_{mn}\,B \end{pmatrix}.$$

$A \otimes B$ is an $(mm')x(nn')$ matrix. For properties of $A \otimes B$ when A and B are square see [7] or [8]. For an mxn matrix C, let $\overline{C}$ denote the (mn)x1 column vector whose components are the elements of C ordered by rows ($\overline{c}_k = c_{ij}$, $k = j + (i-1)n$; $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$). For an mx1 matrix (vector) V, $\overline{V} = V$. With these definitions, it is straightforward to verify that if $C = ANB$ then

$$\overline{C} = (A \otimes B^T)\,\overline{N}.$$

In terms of the above notation, the conditions (5) may be expressed as:

$$F\overline{M} + GV = \overline{B}_0,$$

$$G^T\overline{M} + PV = Y,$$

where

$$F = \sum P_k \otimes X_k X_k^T, \quad G = \sum P_k \otimes X_k, \tag{10}$$

$$B_0 = \sum P_k Z_k X_k^T, \quad Z_0 = \sum P_k Z_k,$$

and

$$P = \sum P_k.$$

Lancaster [6] discusses this notation as well as the direct solution of matrix equations similar to that of (5) for $V = 0$.

Since P is non-singular (the sum of positive definite matrices)

$$V = P^{-1} (Z_0 - G^T \bar{M}),$$

and $\bar{M}$ must satisfy

$$(F - GP^{-1} G^T) \bar{M} = \bar{B}_0 - G P^{-1} Z_0 .$$

The form of these last two equations is similar to that of the simple weight case (8), but the order of the coefficient matrix may be considerably higher.

As with the simple weight case, the translation, linear transformation, and full affine models are separable by a generalization of (9). Explicitly,

$$(F - GP^{-1} G^T)^{-1} = F^{-1} (I - GKG^T F^{-1}),$$

with

$$P + K^{-1} = G^T F^{-1} G,$$

provided the indicated inverses exist.

CONSTRAINTS

In some applications one may have a priori knowledge about the nature of M and desires to restrict or constraint M to be a particular type of matrix. One important application is where M is known to be the matrix of a rotation (orthogonal with determinant of +1) which defines the orientation of a rigid body. Other types of special matrices which will be discussed are: orthogonal ($M^T M = I$), symmetric ($M^T - M = 0$), and skew-symmetric ($M^T + M = 0$).

The definitions of the special matrices above are all expressable as a number of scalar equations of the type $g_{ij}(M) = k_{ij}$ (k_{ij} constant). Thus, all of the constraints may be handled by the method of Lagrange. This requires the minimization of

$$h(M,V) = f(M,V) + g(M),$$

where $f(M, V)$ is as in (2) and

$$g(M) = \sum \lambda_{ij} g_{ij}$$

(summed over all scalar constraint equations). The λ_{ij} are the Lagrangean multipliers to be determined and as the notation implies are considered as elements of an unknown matrix Λ.

The necessary conditions for the constrainted minimum in terms of the gradient defined earlier are now

$$\nabla(M)h = \nabla(M)f + \nabla(M)g = 0,$$

$$\nabla(V)h = \nabla(V)f = 0.$$

The results of performing the operator are then the same as (5) with the terms associated with $\nabla(M)g$ added to the first equation.

The gradient operator can be readily applied to the function $g(M)$ for each of the special matrices being considered. The results are:

Symmetric: $\quad g_{ij} = m_{ji} - m_{ij}, \quad \nabla g = -2\Lambda,$

where

$$\Lambda^T = -\Lambda;$$

Skew-symmetric: $g_{ij} = m_{ji} + m_{ij}, \quad \nabla g = 2\Lambda,$

where

$$\Lambda^T = \Lambda ;$$

Orthogonal: $g_{ij} = \sum_{k=1}^{n} m_{ki}\, m_{kj}, \quad \nabla g = 2 M \Lambda,$

where

$$\Lambda^T = \Lambda .$$

The rotation matrix is the same as the orthogonal case with one additional scalar constraint, namely d(M) = 1 (d(M) denotes determinant of M). Since orthogonal solutions will include the rotations, the rotation case will be considered as a special case of the orthogonal one rather than by adding another Lagrange multiplier.

ORTHOGONAL AND ROTATIONAL SOLUTIONS

The resulting matrix equations for the orthogonal case are:

$$\begin{aligned} \sum P_k M X_k X_k^T + \sum P_k V X_k^T + M \Lambda &= B_0 , \\ \sum P_k M X_k + P V &= Z_0, \\ M^T M &= I , \\ \Lambda^T &= \Lambda , \end{aligned} \tag{11}$$

where B_0, Z_0, and P are defined by (10). M and Λ are square matrices to be determined so that all four matrix equations in (11) are satisfied.

The Equations (11) are quite complex and owing to the non-linearity of two of the equations cannot be handled by any of the techniques used heretofore.

Fortunately, an explicit solution does exist for a simple weighting which allows insight into the geometry of the problem as well as the effects of weighting.

When a single weight is associated to each vector, i.e., $P_k = p_k I$ (p_k positive and I the identity matrix) then the first two equations of (11) can be written as:

$$M(A_0 + \Lambda) + V X_0^T = B_0 ,$$

$$M X_0 + s V = Z_0 ,$$

with A_0, X_0, and s given by (7). Solving the second equation for V and substituting this into the first equation yields

$$M(A + \Lambda) = B, \tag{12}$$

where A and B are defined as they were in (8). From (12) one deduces that

$$B^T B = (A + \Lambda) M^T M(A + \Lambda) = (A + \Lambda)^2$$

since $A + \Lambda$ is symmetric and $M^T M = I$. Letting $H = (A + \Lambda)$ gives

$$\begin{aligned} H^2 &= B^T B, \\ MH &= B. \end{aligned} \tag{13}$$

These last two equations show the dependence of the solution on the square root of the matrix $B^T B$. If H is a non-singular symmetric matrix such that $H^2 = B^T B$, then $\Lambda = H - A$ is symmetric and

$$M = BH^{-1} \tag{14}$$

is orthogonal for $M^T M = H^{-1} B^T B H^{-1} = H^{-1} H^2 H^{-1} = I$. Thus, the M and Λ just defined afford solutions to the Equations (11) when $P_k = p_k I$ for all k.

The conditions imposed on H thus far are not enough to insure uniqueness, e.g., $-H$ is also a non-singular symmetric square root of B^TB if H is. Furthermore, the condition equations in (11) only assure that the extremal values of f are a subset of all solutions. In order to establish the true minimum, the effects of the solutions on the value of f must be examined.

If U is any column vector then $U^TU = tr(UU^T)$. From this, it is straightforward to show that

$$f(M, V) = tr(C) + tr(MAM^T) - 2tr(MB^T)$$

where C is a constant matrix independent of M. Since $tr(MAM^T) = tr(M^TMA) = tr(A)$ when $M^TM = I$, the above expression may be written as:

$$f(M, V) = tr(A + C) - 2tr(MB^T).$$

Hence, the minimum of f is provided by the maximum of $tr(MB^T)$ which is equal to $tr(H)$ for $MH = B$. This establishes that the desired minimum is obtained when H is the symmetric square root of B^TB with largest trace.

Now it is well-known that for an arbitrary square matrix B that B^TB is positive semidefinite. In fact, a classical result of matrix algebra states that any square matrix B can be factored as in (13) with M orthogonal and H a unique positive semidefinite matrix (polar decomposition, see [5] or [8]). From the discussion above, it is then clear that this choice of H provides the desired minimum of the function f for the orthogonal case. If B is singular, however, then any square root of B^TB is also singular and the solution (14) is invalid (the orthogonal part of the polar decomposition is not unique). The construction of the solution for the singular case actually constitutes a proof of the polar

decomposition theorem. However, since the theorem does not cover the rotational case even when B is non-singular, a construction which includes both the singular and rotational cases is given below.

Since $B^T B$ is symmetric and positive semidefinite, there is an orthogonal matrix N such that $N^{-1}B^T BN = D'$ where D' is a diagonal matrix with non-negative entries d'_{ii} ($i = 1, 2, \ldots, m$; m the dimension of B). The ith column of N (denoted as N_i) is a unit eigenvector of $B^T B$ corresponding to the eigenvalue d'_{ii}. Let D be a diagonal matrix such that $D^2 = D'$ ($d^2_{ii} = d'_{ii}$), then the above matrix equation implies

$$(BN_i)^T BN_j = d^2_{ij} = 0, \quad \text{for } i \neq j$$

and

$$|BN_i|^2 = d^2_{ii} .$$

Hence, a complete set of orthonormal vectors Q_i ($i = 1, 2, \ldots, m$) can be constructed so that $BN_i = d_{ii}Q_i$. Let Q denote the orthogonal matrix obtained by juxtoposing the column vectors Q_i in proper order so that $BN = QD$, thus $B = QDN^{-1}$. Now, set

$$M = QN^{-1},$$

$$H = NDN^{-1}, \tag{15}$$

and with the definitions above it is easy to verify that M and H satisfy the Equations (13).* Note also that $Q^{-1}BB^TQ = D^2 = D'$, hence, the d'_{ii} are also eigenvalues of BB^T with eigenvectors Q_i.

*Another polar decomposition is $B = SM$ with $S = QDQ^{-1}$

The trace of H, as given in (15), is just the sum of the entries d_{ii} of D which is a square root of D'. Therefore, the orthogonal least squares solution is obtained by taking all positive square roots in the definition of D. If the resulting M also has a positive determinant then M is also the rotational least squares solution. This will be the case when $d(B) = d(D)d(M) > 0$. If $d(B) < 0$ then the least squares rotation is obtained by changing the sign of the smallest d_{ii} in the definition of D, i.e., all entries of D are positive except the one with least absolute value. When $d(B) = 0$ (B singular) the constructed M (using non-negative d_{ii}) may or may not have a positive determinant. Should the determinant of M be negative then changing the sign of any vector Q_i (the ith column of Q) corresponding to a zero d_{ii} will change the sign of d(M) without changing the value of $tr(H) = tr(D)$. The orthogonal solution is unique provided $d(B) \neq 0$. The rotational solution is unique unless the smallest eigenvalue is a multiple root and $d(B) \leq 0$.

There are many interesting interpretations of the matrix M just constructed. It is the orthogonal or rotation matrix which: (1) Minimizes Equation (3) for the assumed weighting, (2) Maximizes $tr(MB^T)$ thus minimizes the norm of B - M, (3) Provides a polar decomposition of B, and (4) Is a solution of the matrix equation $MB^T - (MB^T)^T = 0$. This last equation stems from the decomposition $B = SM$, its significance will be discussed later. It is also noteworthy that the orthogonal or rotational M is completely independent of the matrix A whereas the unconstrainted solution was highly dependent upon A.

ATTITUDE DETERMINATION

Since the advent of the space age a wide variety of problems have been pursued within the label of attitude determination. These include determination

of: (1) The orientation of all three coordinate axes of a stabilized vehicle, (2) The direction and/or rate of the spin axis of a spin-stabilized vehicle, and (3) Either of the above as a function of time. Each of these problems has been solved for a wide range of sensors. In addition, many parametrizations of the rotation group have been employed as the independent parameters to be determined. The weighted least squares rotation matrix of the previous section provides a solution to many of these problems without re-formulating equations for each new sensor. It is applicable to any sensor whose output can be formed into a vector of components relative to the coordinate system whose "attitude" is to be determined. The desired parameters can then be obtained from the determined matrix.

If the model assumes that the attitude is being held by inertial sensors then readings at different times may be combined. This technique has been used successfully for the Orbiting Astronomical Observatory [2] to the extent of determining attitude from magnetometer data only while in darkness (after the magnetometers were aligned by techniques herein). When the attitude is a function of time and M_1 is the least squares attitude matrix at time t_1 and M_2 the least squares matrix at t_2 then $\Delta M = M_2 M_1^{-1}$ defines the spin axis and angle of rotation during the interval $t_2 - t_1$. If the spin axis and rate are considered constant and the data taken at equal time intervals; a least squares estimate of the rate matrix may be obtained by denoting the measured vectors at t_i as X's and those taken at t_{i+1} as Z's.

The usual statement of the attitude problem is without the translation since position and orientation data are provided by different sensors. For this case,

the second equation of (11) is omitted and V set to zero. This still leads to equations of the form (13), but with a simplified B matrix, i.e.,

$$B = B_0 = \sum p_k Z_k X_k^T .$$

Since p_k is assumed positive, let $p_k = q_k^2$; the matrix B may then be written as

$$B = \sum W_k V_k^T ,$$

with $V_k = q_k X_k$ and $W_k = q_k Z_k$. With this notation

$$\mathrm{tr}(M B^T) = \mathrm{tr}\left(\sum M V_k W_k^T\right) = \sum W_k^T M V_k = \sum W_k \cdot (M V_k) \qquad (16)$$

As was mentioned earlier, every M satisfying (13) is a solution of the matrix equation

$$M B^T - (M B^T)^T = 0 . \qquad (17)$$

Thus, the desired solution is among the solutions of (17). For the present case ($B = B_0$) this yields

$$\sum (M V_k) W_k^T - W_k (M V_k)^T = 0 . \qquad (18)$$

The left-hand side of the above equation is a skew-symmetric matrix, thus represents only $n(n-1)/2$ independent scalar equations. In three-space, such a matrix is isomorphic to a 3x1 vector. Let $\tilde{U}$ denote the skew-symmetric matrix formed from the vector U such that for any vector V, $\tilde{U}V = U \times V$. Then the independent scalar equations given by (18) can be expressed as:

$$\sum W_k \times (M V_k) = 0 . \qquad (19)$$

In summary, the desired weighted least squares attitude matrix is that rotation matrix which satisfies (19) and maximizes (16).

Expressing the condition equations in terms of the cross and dot product provides a physical or geometric interpretation of the solution which leads to some interesting observations. The simplest case is when only one measured vector is available. In this case the answer is not unique, but a rotation about the line $W_1 \times V_1$ by the angle between W_1 and V_1 provides the "shortest" rotation [2]. Note that an error in the length of Z_1 does not effect the answer.

A more practical situation and perhaps the one that has received the most attention is when two measurements are given. Equation (19) requires that the plane defined by V_1 and V_2 (plane I) be rotated into the plane defined by W_1 and W_2 (plane II) such that

$$|W_1 \times (M V_1)| = |(M V_2) \times W_2| .$$

This last condition requires the area of the triangle formed by W_1, MV_1, and $W_1 - MV_1$ (triangle I) to be equal to the area of the triangle formed by W_2, MV_2, and $W_2 - MV_2$ (triangle II). Since M preserves length ($|MV| = |V|$) this requirement may be stated as:

$$|W_1| \; |V_1| \; \sin\theta_1 = |W_2| \; |V_2| \; \sin\theta_2 ,$$

or

$$p_1 \; |Z_1| \; |X_1| \; \sin\theta_1 = p_2 \; |Z_2| \; |X_2| \; \sin\theta_2 , \tag{20}$$

where θ_i is the positive angle between MX_i and Z_i, $i = 1, 2$.

The requirement that plane I be rotated into plane II may be satisfied as follows: Let N_1 and N_2 be orthogonal vectors of unit length in plane I and $N_3 = N_1 \times N_2$. Similarly, let Q_1 and Q_2 be orthogonal unit vectors in plane II and $Q_3 = Q_1 \times Q_2$. $N = (N_1, N_2, N_3)$ and $Q = (Q_1, Q_2, Q_3)$ are then rotation matrices and the matrix M such that $MN = Q$ or $M = QN^{-1}$ is a rotation matrix which indeed satisfies the requirement. Note that this form of M is precisely that of the polar decomposition solution given by (15). It remains to define N_1, N_2, Q_1, and Q_2 explicitly so that all other conditions are satisfied.

It is geometrically obvious and a routine matter to show that when $|V_1| = |V_2|$ and $|W_2| = |W_2|$ the correct solution is given by:

$$N_1' = V_1 - V_2, \quad N_2' = V_1 + V_2, \quad N_i = N_i' / |N_i'|, \quad (i = 1, 2)$$

$$Q_1' = W_1 - W_2, \quad Q_2' = W_1 + W_2, \quad Q_i = Q_i' / |Q_i'| .$$

This suggests a general solution of the form $N_1' = xV_1 - V_2$ and $N_2' = yV_1 + V_2$ where x and y are scalars of proportionally depending on the relative lengths of the given vectors. Indeed, it can be shown that, if

$$V_1 \cdot V_1 xy + V_1 \cdot V_2 (x - y) - V_2 \cdot V_2 = 0,$$

and (21)

$$W_2 \cdot W_2 xy - W_1 \cdot W_2 (x - y) - W_1 \cdot W_1 = 0,$$

then N_1' and N_2' are eigenvectors of B^TB (polar decomposition). The first equation of (21) is just the requirement that N_1' and N_2' be orthogonal. The vectors BN_1' and BN_2' are proportional to $Q_1' = W_1 - yW_2$ and $Q_2' = W_1 + xW_2$ respectively, i.e., Q_1' and Q_2' are eigenvectors of BB^T and the second equation of (21) is the condition for their orthogonality.

Solving the simultaneous Equations (21) yields

$$x - y = a/c,$$

$$xy = b/c,$$

with

$$a = (V_2 \cdot V_2)(W_2 \cdot W_2) - (V_1 \cdot V_1)(W_1 \cdot W_1), \quad b = (W_1 \cdot W_1)(V_1 \cdot V_2) + (V_2 \cdot V_2)(W_1 \cdot W_2),$$

$$c = (V_1 \cdot V_1)(W_1 \cdot W_2) + (W_2 \cdot W_2)(V_1 \cdot V_2).$$

These last two equations have two solutions, given by

$$x = \frac{a \pm \sqrt{a^2 + 4bc}}{2c}, \quad y = \frac{-a \pm \sqrt{a^2 + 4bc}}{2c},$$

where the signs of the radicals must be consistent. Both solutions will also satisfy (19) if

$$N_1' = xV_1 - V_2, \quad N_2' = yV_1 + V_2, \quad N_i = N_i'/|N_i'|, \quad (i = 1, 2)$$

$$Q_1' = W_1 - yW_2, \quad Q_2' = W_1 + xW_2, \quad Q_i = Q_i'/|Q_i'|,$$

$$N_3 = N_1 \times N_2, \qquad Q_3 = Q_1 \times Q_2,$$

$$N = (N_1, N_2, N_3), \qquad Q = (Q_1, Q_2, Q_3),$$

and

$$M = QN^{-1}.$$

The solution that maximizes (16) may be ascertained by expressing V_1, V_2, W_1, and W_2 in terms of N_1', N_2', Q_1', and Q_2'. When this is done, it is found that x and y must also satisfy the condition $x + y > 0$. The sign of the radical in the definitions of x and y can be selected so that this condition is met. This, then,

completes the solution when two measured vectors are given. This solution appears in [2] without proof and Fraiture [4] offers a different construction.

Perhaps, the most important aspect of the two measurement solution is the insight it provides to the effects of weighting. The length of each of the vectors W_1 and W_2 obviously affects the solution, and their lengths are a function of the weights as well as the lengths of the measured vectors Z_1 and Z_2. Therefore, an error in the length of either Z_1 or Z_2 has the same effect as a weight factor and biases the solution. This implies that the length of the measurement Z_i should be made equal to that of X_i. One arrives at the same conclusion by arguing that, since the rotation cannot change the length, any deviation in length is due to measurement noise (provided any misalignment has been eliminated) and should be removed. Likewise, the lengths of X_1 and X_2 may bias the solution if they differ. This is particularly true when a sensor measuring only direction is combined with one measuring length as well. Thus, it is concluded, that to obtain an unbias attitude (rotation) matrix all data should be normalized so that the only length appearing in the V's and W's is that due to the weighting factors. This same conclusion is reached in [2], but by a completely different argument.

When more than two measurements are given the polar decomposition of the previous section provides a solution. The solutions in [3] also depend upon the polar decomposition in slightly different form. [1] offers two solutions quite different in nature. The polar decomposition solution indicates that care should be taken when employing an iterate or differential correction type of solution. If the matrix B is non-singular then there are four solutions to the condition equations, all satisfying the constraints: assume $d(B) > 0$, then the desired solution is with all three positive square roots, however, any combination

of two negative and one positive entries in D also yields a rotation matrix. A similar argument exists for $d(B) < 0$. These spurious solutions could cause false convergence with a poor initial estimate.

SYMMETRIC AND SKEW-SYMMETRIC SOLUTIONS

The constrainted equations for the symmetric and skew-symmetric cases are very similar, differing only by a plus or minus sign. Because of this, the solutions are also similar and will be treated together. The resulting equations are still linear in M and Λ; thus, the general case can be handled by reconstructing the independent scalar equations into a single matrix equation.

As with the previous cases discussed, the weighting $P_k = p_k I$ offers a simplified solution. In this case, the equations to be solved are:

$$M A = B \pm \Lambda,$$

$$M^T = \pm M,$$

$$\Lambda^T = \mp \Lambda.$$

The upper signs pertain to the symmetric case, whereas the lower signs denote the skew-symmetric case. The matrices A and B are as previously defined. Once M has been determined, V is given as in (8).

From the above equations, one deduces that

$$B \pm B^T = M A \pm (M A)^T = M A + A M. \tag{22}$$

Since A is symmetric, there exists an orthogonal matrix Q such that $Q^{-1} A Q = D$ is a diagonal matrix. Equation (22) is then equivalent to:

$$M'D + DM' = Q^{-1}(B \pm B^T)Q$$

with $M' = Q^{-1}MQ$. The component equations are

$$(d_{ii} + d_{jj})\, m'_{ij} = s_{ij}, \quad (i, j = 1, 2, \ldots, n)$$

where $S = Q^{-1}(B \pm B^T)\, Q$ is symmetric or skew-symmetric depending on the case being considered. Therefore, M' is symmetric or skew-symmetric respectively and it follows that $M = Q M' Q^{-1}$ is also.

REFERENCES

[1] Paul B. Davenport, "A Vector Approach to the Algebra of Rotations With Applications," NASA TN D-4696, Aug. 1968; or GSFC X-542-69-417, Nov. 1965.

[2] Paul B. Davenport, "The Attitude Determination System for the Orbiting Astronomical Observatory," A.F. Report No. SAMSO-TR-69-417, Vol. 1, Proceedings of the Symposium on Spacecraft Attitude Determination, Oct. 1969, pp. 249-256.

[3] J. L. Farrell, et al., "A Least Squares Estimate of Satellite Attitude," SIAM Review, 8 (1966), pp. 384-386.

[4] Luc Fraiture, "A Least-Squares Estimate of the Attitude of a Satellite," J. Spacecraft and Rockets, 7 (1970), pp. 619-620.

[5] F. R. Gantmacher, "The Theory of Matrices," Vol. 1, Chelsea, New York, 1959.

[6] Peter Lancaster, "Explicit Solutions of Linear Matrix Equations," SIAM Review, 12 (1970), pp. 544-566.

[7] C. C. MacDuffee, "The Theory of Matrices," Chelsea, New York, 1956, pp. 81-97.

[8] Marvin Marcus, "Basic Theorems In Matrix Theory," Nat. Bureau of Stds., Appl. Math. Series No. 57, 1960.

[9] Charles M. Price, "The Matrix Pseudoinverse and Minimal Variance Estimates," SIAM Review, 6 (1964), pp. 115-123.

BIBLIOGRAPHY

Matrix Algebra

1. Birkhoff and MacLane, "A Survey of Modern Algebra," Macmillan, New York, 1956.

2. Thrall and Tornheim, "Vector Spaces and Matrices," Wiley, New York, 1957.

Calculus

3. Angus E. Taylor, "Advanced Calculus," Ginn, New York, 1955, chap. VI.

General

4. Korn and Korn, "Mathematical Handbook for Scientists and Engineers," McGraw-Hill, New York, 1968.

5. W. F. Freiberger, et al., "The International Dictionary of Applied Mathematics," Van Nostrand, Princeton, 1960.

Alignment Models

6. R. des Jardins, "In-Orbit Startracker Misalignment Estimation on the OAO," same as [2], pp. 143-153.

7. R. des Jardins, "MESS (Misalignment Estimation Software System for In-Flight Alignment and Calibration of Spacecraft Attitude Sensors)," AAS No. 71-357, AAS/AIAA Astrodynamics Specialists conf. 1971, Ft. Lauderdale, Fla., Aug. 1971.

8. F. G. Unger, "Vector and Matrix Representations of Inertial Instruments," 11th Annual East Coast Conf. on Aerospace and Navigational Elect., Baltimore, Md., Oct. 1964, pp. 1.1.2-1.1.2-8.

Rotations

9. Paul B. Davenport, "Mathematical Analysis For The Orientation and Control Of The Orbiting Astronomical Observatory Satellite," NASA TN D-1668, 1963.

10. Paul B. Davenport, "Slewing About Non-Orthogonal Axes," AAS No. 71-389, AAS/AIAA Astrodynamics Specialists Conf. 1971, Ft. Lauderdale, Fla., Aug. 1971.

11. R. A. Frazer, W. J. Duncan, and A. R. Collar, "Elementary Matrices," University Press, Cambridge, 1957, pp. 246-256.

12. J. W. Gibbs, "Vector Analysis," Yale University Press, New Haven, 1958, pp. 332-371.

13. Carl Grubin, "Vector Representation of Rigid Body Rotation," Amer., J. Phys., 30 (1962), pp. 416-417.

14. G. A. Korn and T. M. Korn, "Mathematical Handbook for Scientists and Engineers," McGraw-Hill, New York, 1968, pp. 471-479.

15. Arthur Mayer, "Rotations and Their Algebra," SIAM Review, 2 (1960), pp. 77-122.

16. R. E. Mortensen, "On Systems for Automatic Control of the Rotation of a Rigid Body," Elect. Res. Lab., U. of Calif., Series No. 63, Issue No. 23, Nov. 1963.

17. Peter Rastall, "Rotations and Lorentz Transformations," Nuclear Phys., 57 (1964), pp. 191-199.

18. John Stuelpnagel, "On The Parametrization of the Three-Dimension Rotation Group," SIAM Review, 6 (1964), pp. 422-430.

19. J. L. Synge, "On Classical Dynamics," Encyclopedia of Physics, Vol. III/1, pp. 17-25.

20. E. T. Whittaker, "A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, Dover Publications, New York, 1944, pp. 8-13.

Attitude Determination

21. P. A. Bracken and L. R. Davis, "A Method Used For The Determination Of The Explore 26 Spin Axis Position," NASA-TM-X-63724, Oct. 1969.

22. A. E. Bryson and W. Kortüm, "Estimation of the Local Attitude of Orbiting Spacecraft," International Fed. of Auto. Control, International Conf., 3rd, Toulouse, France, Mar. 1970.

23. S. Chris Dunker and Alva M. Butcher, "Prelaunch Analysis and Backup Attitude Determination Plan For The Satellite San Marco-C," GSFC-NASA, X-540-70-68, Jan. 1970.

24. Ai Chun Fang, "Delta Pac Attitude Determination and Error Prediction," GSFC-NASA, X-732-70-338, Sept. 1970.

25. Robert Fischl, "Analysis of the Effect of Observation Errors on the Estimation of Satellite Attitude," Drexel Univer., Philadelphia.

26. Edwin C. Foudriat, "Analysis of Limited Memory Estimators and Their Application to Spacecraft Attitude Determination," NASA-CR-1654, Aug. 1970.

27. G. S. Goodman and G. Margulies, "Attitude Determination of an Orbiting Satellite," Western Development Lab., Philco, WDL-TN62-7, Nov. 1962.

28. Pranas A. Grusas, "Satellite Attitude Determination from Celestial Sightings," Journal of Spacecraft and Rockets, Vol. 6 (1969), pp. 1007-1012.

29. Harold A. Hamer, "Manual Procedure For Determining Position In Space From Onboard Optical Measurements," NASA TN D-1852, Dec. 1964.

30. Raul R. Hunziker and Leo B. Schlegel, "San Marco Satellite Attitude Determination Based On Aerodynamic Drag Measurements," International Astronautical Federation, International Astronautical Congrees, 21st, Konstanz, West Germany, Oct. 1970.

31. Charles F. Miller, Jr., "A Graphic Method For Determining The Absolute Attitude Of Sounding Rocket Vehicles," NASA TN D-5172, May 1969.

32. A. J. Mooers, et al., "Breadboard Design of A Scanning Celestial Attitude Determination System," Control Data Project No. 9563, RD 2003, Minneapolis, Minn., Nov. 1966.

33. J. Spencer Rockefort, et al., "Attitude Determination and Data Transmission Systems for Space Vehicles," AFCRL-69-0146, USAF, Bedford, Mass., Oct. 1968.

34. G. M. Ross and D. A. King, "Radio Astronomy Explorer Attitude Determination System - Vol. 5 Deterministic Three-Axis Attitude Determination Program," IBM, Federal Systems Division, Gaithersburg, Md., May 1969.

35. Grace Wahba, "Problem 65-1, A Least Squares Estimate of Satellite Attitude," SIAM Review, 7 (1965), p. 409.

36. R. D. Werking, et al., "SAS-A Attitude Control Prelaunch Analysis and Operation Plan," GSFC-NASA, X-541-70-400, Nov. 1970.